SkiLD: Unsupervised Skill Discovery Guided by Factor Interactions

Zizhao Wang*, Jiaheng Hu*, Caleb Chuck*, Stephen Chen,

Roberto Martín-Martín, Amy Zhang, Scott Niekum, Peter Stone

University of Texas at Austin, Sony AI

Background: Unsupervised Skill Discovery

Learn reusable behaviors through reward-free interaction

- Represented as a policy $\pi(a|s, z)$ conditioned on the skill z.

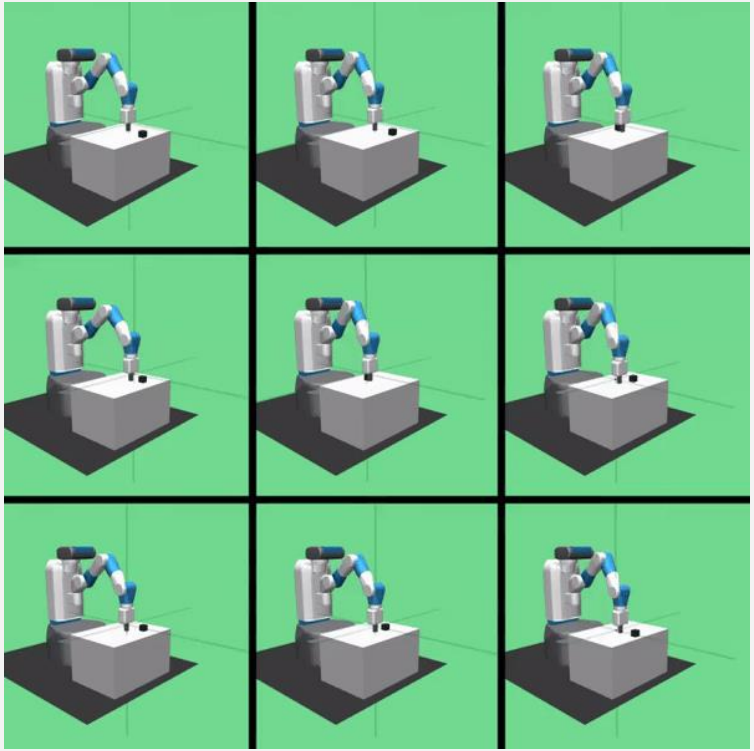
Key idea: visit distinguishable states for each z.

Eysenbach, Benjamin, et al. "Diversity is all you need: Learning skills without a reward function."

Background: Unsupervised Skill Discovery

Drawbacks:

- Skills are not semantically meaningful.
- Do not scale to environments with multiple state factors (e.g., objects).

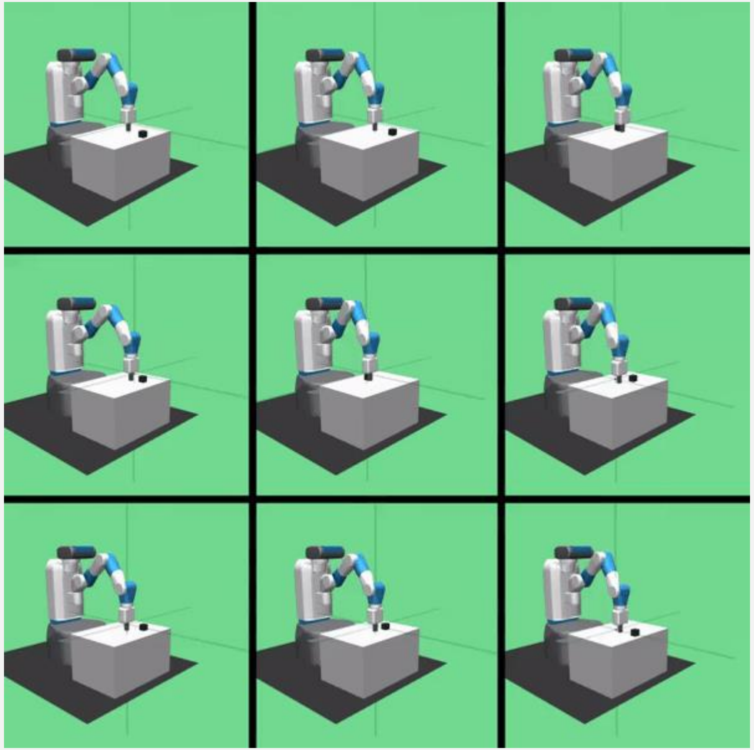


https://seohong.me/projects/csd/

Motivation

Drawback Causes:

- As state space grows exponentially w.r.t. state factors, it is challenging to reach all possible states.
- In addition, many states are not very meaningful for downstream tasks.



Motivation

Another way to represent skills: interactions

- Many skills can be described as interactions between state factors.
  - For example, driving a nail is an interaction between a human, a hammer and a nail.
  - Factor interactions are naturally semantically meaningful.
  - For many downstream tasks, inducing interactions is more useful than visiting some random states.
- Moreover, # inducible interactions << # reachable states.
  - Hence, it is easier to learn skills to cover the interaction space.

Method: skill representation

We propose that, for multi-factor state environments,

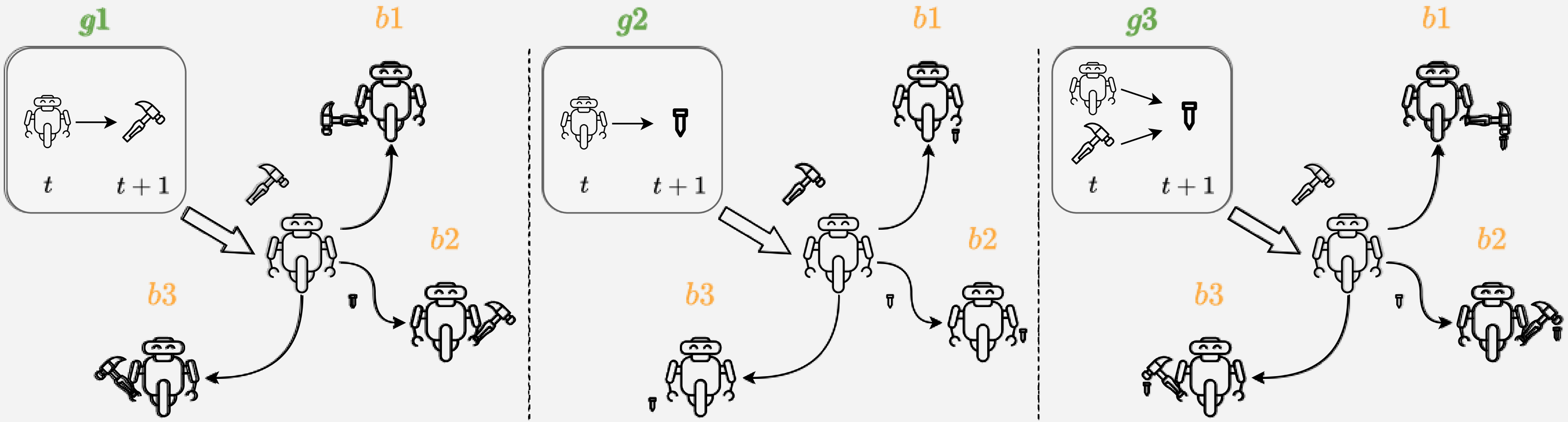
skill discovery should learn to induce diverse interactions between state factors.

Method: skill representation

skill discovery should learn to induce diverse interactions between state factors.

$$z = (g, b)$$

desired interactions — diversity variable

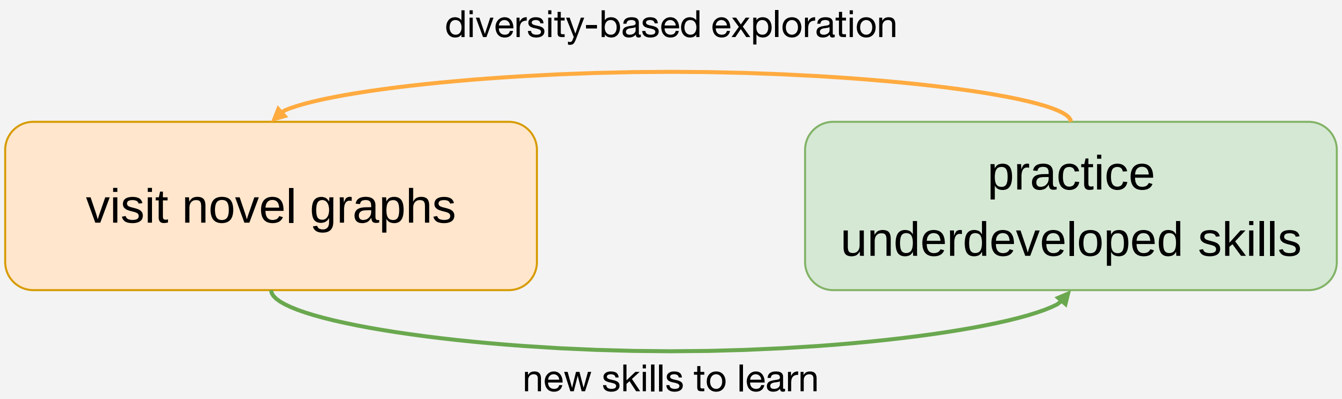


Method: high-level graph selection policy $\pi_{\mathcal{G}}(g|s)$

Graph selection policy

$$\pi_{\mathcal{G}} : \mathcal{S} \rightarrow \mathcal{G}$$

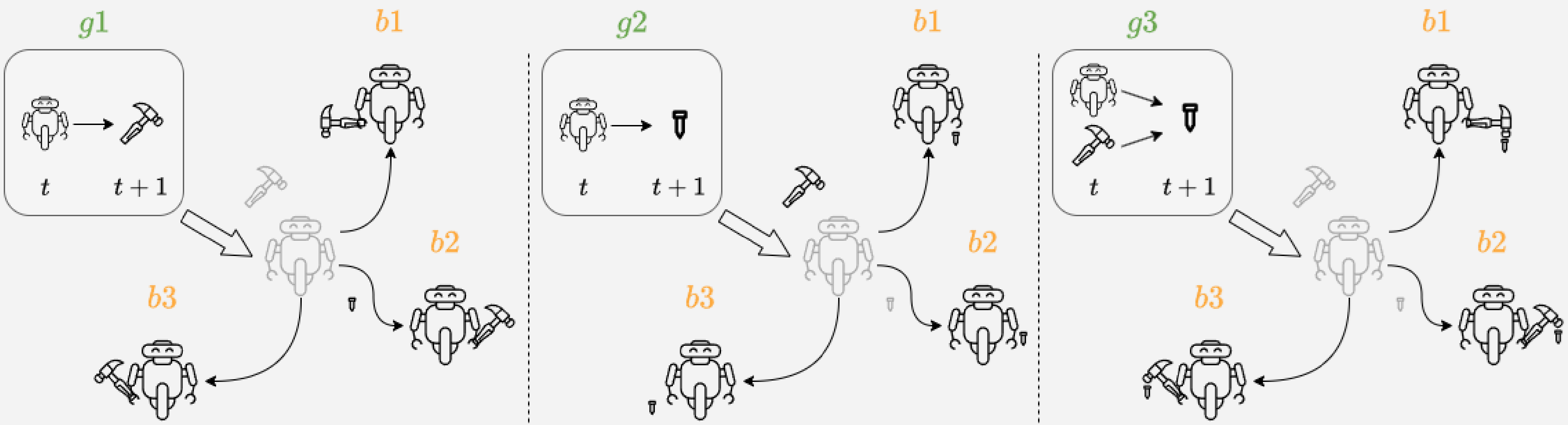
It guides the exploration and training of skill policy



Method: low-level skill policy $\pi_{\text{skill}}(a|s,z)$

The skill policy $\pi_{\text{skill}}(a|s,z)$, where $z = (g, b)$, has two goals:

- achieve the desired graph g
- then, further change s to diverse values while maintaining the graph



SkiLD: Unsupervised Skill Discovery Guided by Factor Interactions

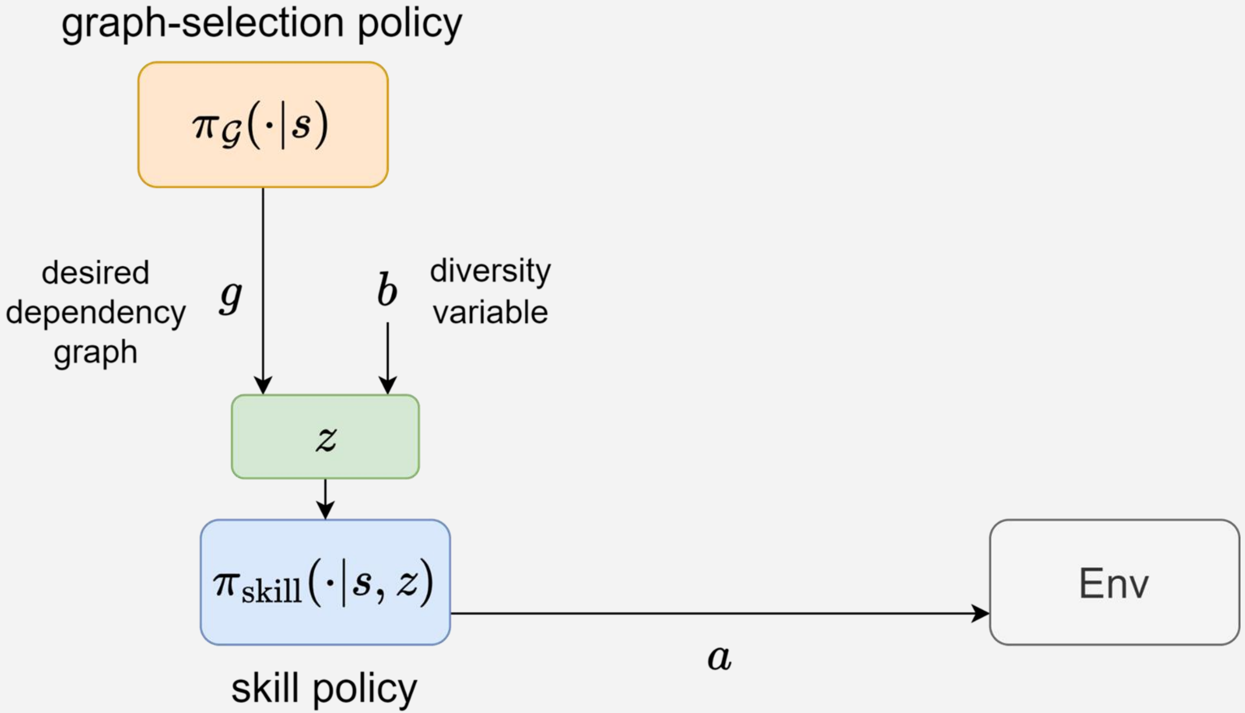
Zizhao Wang*, Jiaheng Hu*, Caleb Chuck*, Stephen Chen, Roberto Martín-Martín, Amy Zhang, Scott Niekum, Peter Stone

University of Texas at Austin, Sony AI

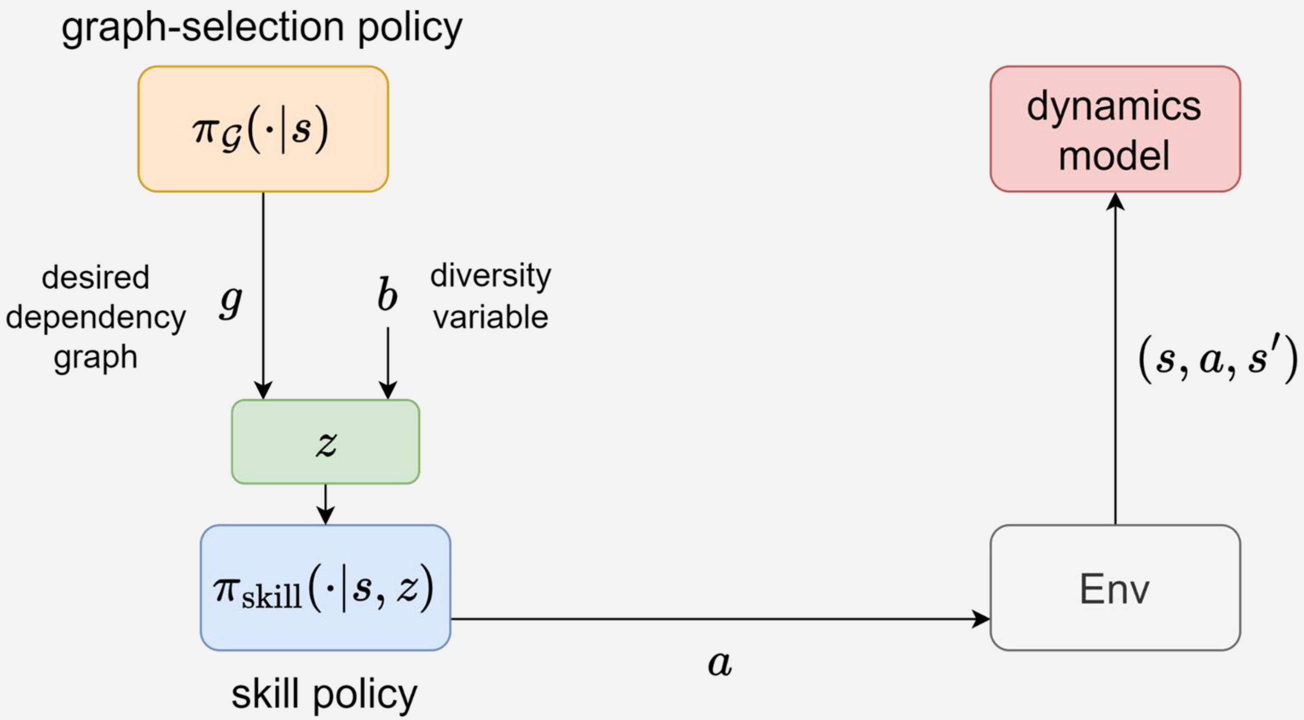
Project Page

Paper

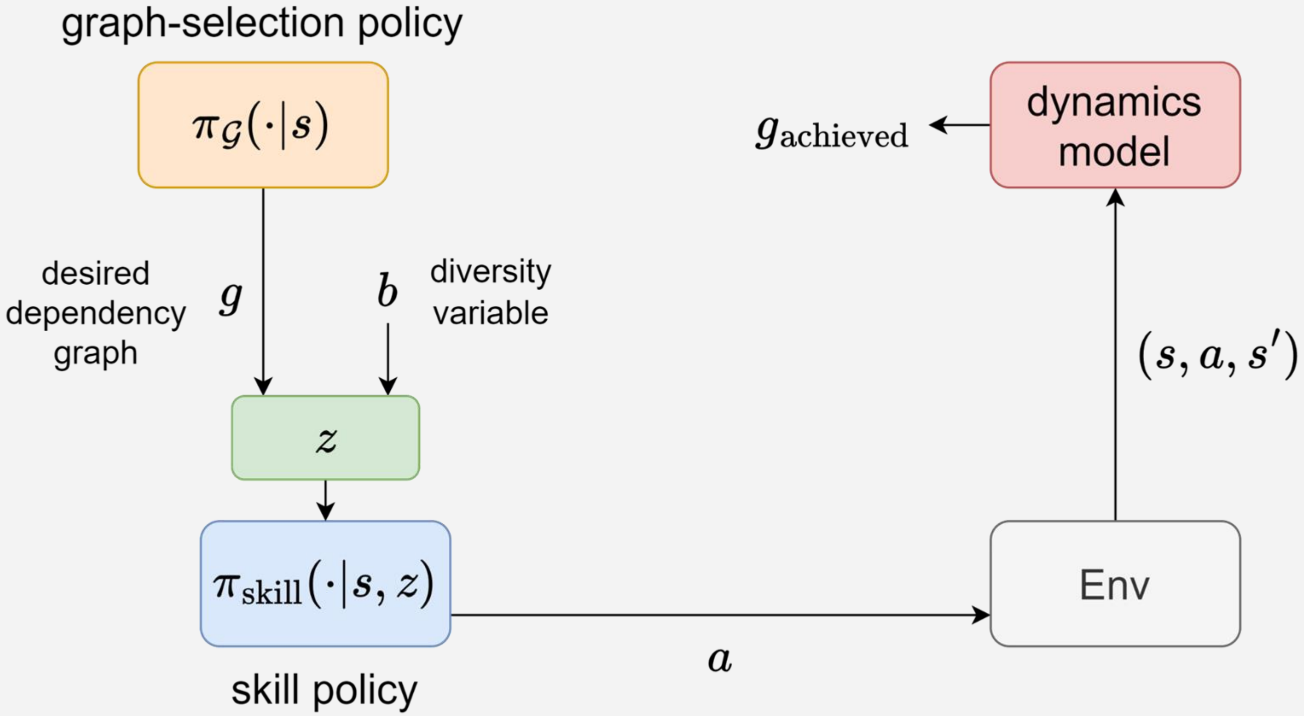
Method: pipeline - skill learning



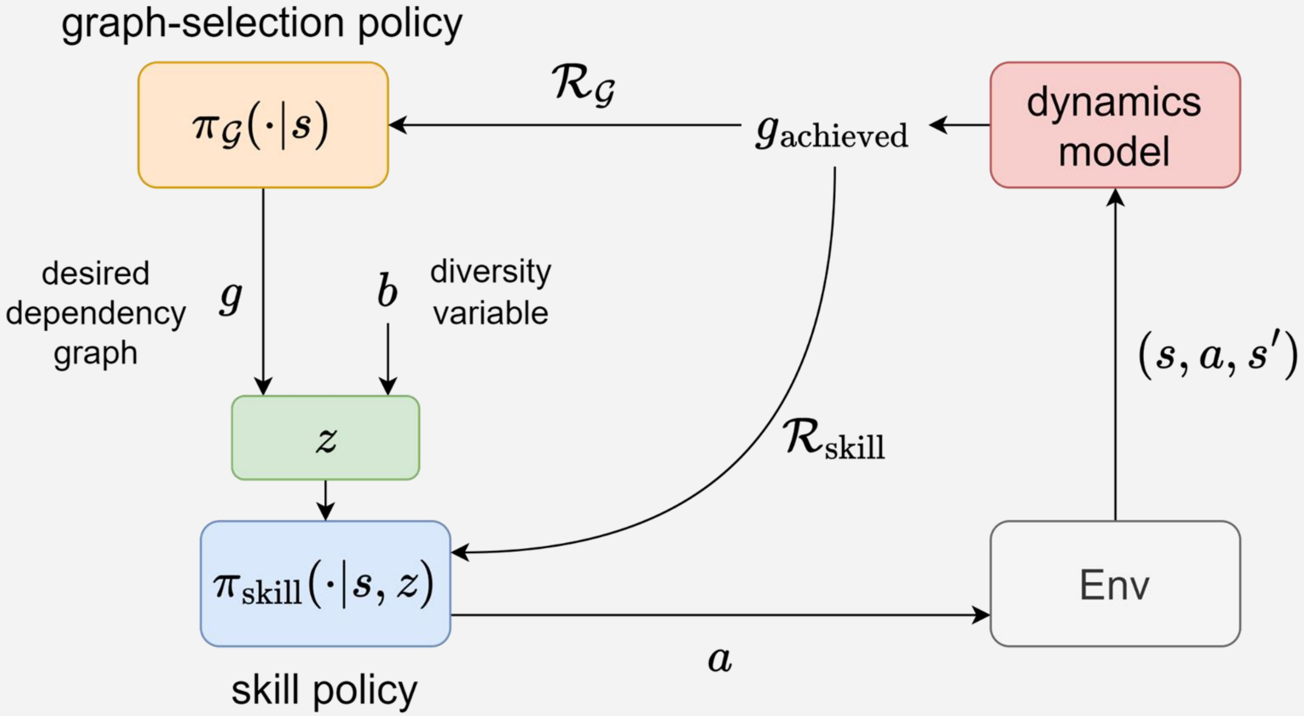
Method: pipeline - skill learning



Method: pipeline - skill learning



Method: pipeline - skill learning



Method: pipeline - task learning